

On Secrecy Capacity of Minimum Storage Regenerating Codes ^{*}

Kun Huang¹, Udaya Parampalli², and Ming Xian¹

Abstract. In this paper, we revisit the problem of characterizing the secrecy capacity of minimum storage regenerating (MSR) codes under the passive (l_1, l_2) -eavesdropper model, where the eavesdropper has access to data stored on l_1 nodes and the repair data for an additional l_2 nodes. We study it from the information-theoretic perspective. First, some general properties of MSR codes as well as a simple and generally applicable upper bound on secrecy capacity are given. Second, a new concept of *stable* MSR codes is introduced, where the stable property is shown to be closely linked with secrecy capacity. Finally, a comprehensive and explicit result on secrecy capacity in the linear MSR scenario is present, which generalizes all related works in the literature and also predicts certain results for some unexplored linear MSR codes.

Key Words: MSR Codes, Repair Data, Secrecy Capacity, Upper Bounds.

1 INTRODUCTION

Distributed storage systems (DSSs) are an essential part of large scale data storage systems required for many new emerging distributed networking applications such as social networking, video sharing, peer to peer networking and large scale data centres. As is common in such storage systems, redundancy is indispensably introduced to ensure reliability and availability owing to frequent node failures. The main approaches to introduce redundancy in DSSs are through replication, erasure codes, and more recently using regenerating codes [7]. Erasure codes in general can achieve higher reliability for the same level of redundancy when compared to the schemes that provide replication [6]. Regenerating codes are a recent innovation of erasure codes that has efficient performance on repair of failed nodes in DSSs [8].

1.1 Regenerating Codes.

Regenerating codes [7] are a family of maximal distance separable (MDS) codes determined by a tradeoff between the amount of storage per node and the repair bandwidth. In the framework of regenerating codes, an encoded data file is split into $n\alpha$ symbols and then dispersed across n nodes, where all the symbols are drawn from a finite field \mathbb{F}_q and each node stores a collection of α symbols. The dispersing manner requires that any data collector can retrieve the original data message by connecting to any k out of n nodes. The node repair can be accomplished by permitting a new node to connect to any d helper nodes from the surviving $(n - 1)$ nodes by downloading $\beta \leq \alpha$ symbols from each node. In the literature, a regenerating code is represented by a parameter set $\{n, k, d, \alpha, \beta, B\}$, where B is the size of original data message and $d\beta$ is the total amount of data transferred for node repair that is termed repair bandwidth.

The cut-set bound based on the concept of information flow [4] requires that the parameters of a regenerating code must necessarily satisfy:

$$B \leq \sum_{i=1}^k \min\{\alpha, (d - i + 1)\beta\}. \quad (1)$$

^{*} ¹Kun Huang and Ming Xian are with State Key Laboratory of Complex Electromagnetic Environment Effects on Electronics and Information System, National University of Defense Technology, Changsha, 410073, China (khuangresearch923@gmail.com; qwertmingx@tom.com).

² Udaya Parampalli is with Department of Computing and Information Systems, University of Melbourne, VIC 3010, Australia (udaya@unimelb.edu.au).

In [7], Dimakis et al derive the above tradeoff between the per node storage α at each node and repair bandwidth $d\beta$. The codes that can achieve this tradeoff curve are called *optimal* regenerating codes. Two extreme points on this tradeoff curve are of particular concern, namely, minimum bandwidth regenerating (MBR) point and minimum storage regenerating (MSR) point, respectively representing codes with the least repair bandwidth and ones with the least per node storage. As shown in [7], the parameters of MBR and MSR codes are given by:

$$\begin{cases} (\alpha_{\text{MSR}}, \beta_{\text{MSR}}) = (\frac{B}{k}, \frac{B}{k(d-k+1)}) \\ (\alpha_{\text{MBR}}, \beta_{\text{MBR}}) = (\frac{2dB}{k(2d-k+1)}, \frac{2B}{k(2d-k+1)}) \end{cases} \quad (2)$$

In the literature, three repair models are considered: functional repair, exact repair, and exact repair of systematic nodes [8]. Exact repair can regenerate the exact replicas of the lost data in the failed nodes and thus is preferred in practical systems [5]. In the exact repair scenario, Shah et al in [9] demonstrate that most interior points on the storage-bandwidth tradeoff curve are not achievable. For those possibly reachable interior points, constructions of codes are rare [12,13]. In addition, Duursma in [10,11] derive some new outer bounds for regenerating codes with exact repair.

Up to now, several constructions with the exact repair property for MBR and MSR codes have been proposed. In [15], Rashmi et al employ product matrix to construct MBR codes for all parameters and MSR codes with $\{d \geq 2k - 2\}$. In the MSR scenario, significant progress have been made. From the overall perspective, there are two classes of MSR codes, i.e., the scalar MSR codes with $\{\beta = 1\}$ [15,16,17,18,19,20] and vector MSR codes with $\{\beta = (n-k)^x\}$ where $x \geq 1$ [21,22,23,24,25,26,27,28,29]. Many of these constructions are established on interference alignment. As explained in [20], interference alignment is the necessity of constructing linear scalar MSR codes and these linear scalar MSR codes only exist when $d \geq 2k - 2$. From another point of view, this existing restriction exactly corresponds to the low rate regime, i.e., $\frac{k}{n} \leq \frac{1}{2n} + \frac{1}{2}$. As for the high rate codes with $\{\frac{k}{n} > \frac{1}{2}\}$, vector MSR codes are available as they are free from the parameter constraints of (n, k) . However, many of these vector codes only allow efficient repair of systematic nodes [23,24,25,26,27,28,29], such as Zigzag codes [23]. In [21,22], the authors present vector MSR codes allowing efficient repair for parity nodes as well, where the code given in [21] is a variant of Zigzag code.

In addition to repair efficiency, there are many other design features required by DSSs such as security [30,31,32,33,34,35], local-repairability [33,38,39,40], optimality of updating [23,28,29], etc. Our concern in this paper is on securing DSSs against eavesdroppers attempting to obtain any knowledge of the original data.

1.2 Secure Regenerating Codes.

Since the nodes of DSSs are widely spread across the network, individual nodes may be compromised and as a result the data stored is vulnerable to eavesdropping. There are mainly two kinds of attacker models considered in the literature: passive eavesdropper model and active eavesdropper model [3]. Compared to the former, active eavesdropper can modify the data or even inject new data into the compromised nodes. Our eavesdropper model considered in this paper is the passive one as given in [31]. In this model, eavesdropper has access to the data stored on l_1 nodes as well as the repair data for an additional l_2 nodes. Here, we only consider the situation of exact repair¹.

Related work: The issue of designing secure regenerating codes against eavesdropping was firstly addressed in [30] and [31]. The authors in [30] considered the initial setting that an eavesdropper observes the contents of $l < k$ nodes of the storage system, and analyzed the regenerating code's secrecy capacity

¹ Functional repair scheme requires ceaselessly updating the data stored in nodes undergoing repair, which may leak substantial linear combinations of data to eavesdroppers and enable the eavesdroppers to retrieve the original data just by solving the linear equations. This is another reason why exact repair is superior to functional repair.

(i.e., the maximal file size that can be securely stored). An upper bound of the secrecy capacity and a secure MBR code that can attain this bound are proposed in [30]. Extending the initial eavesdropper setting [30], authors in [31] modeled the eavesdropper as one obtaining access to the data stored on l_1 nodes as well as the repair data for an additional l_2 nodes, with $l_1 + l_2 < k$. The secure product-matrix-based MBR coding scheme proposed in [31] can achieve the bound derived in [30] with $l = l_1 + l_2$. Achievability of the bound for secure product-matrix-based MBR codes in [31] can be attributed to the fact that the repair bandwidth $d\beta$ equals to per node storage α in the MBR scenario. In other words, the (l_1, l_2) -eavesdropper cannot obtain any extra information other than the contents of $l = l_1 + l_2$ nodes in the MBR scenario. Hence, under the (l_1, l_2) -eavesdropper model, the upper bound in [30] still holds for the secure file size $B^{(s)}$ of MBR codes:

$$B^{(s)} \leq \sum_{i=l+1}^k \min\{\alpha, (d-i+1)\beta\}, \quad (3)$$

where $l = l_1 + l_2$. Authors in [31] further considered the design of secure MSR codes based on product-matrix codes, but the secure MSR coding scheme is only capable of storing $(k - l_1 - l_2)(\alpha - l_2\beta)$ -sized secure files, which reaches the bound (3) only when $l_2 = 0$. The intuition here indicates that the (l_1, l_2) -eavesdropper can obtain more information than the contents of $(l_1 + l_2)$ nodes in the MSR scenario, as the repair bandwidth $d\beta$ is larger than $\alpha = (d - k + 1)\beta$ that is the amount of data stored on each of those l_2 nodes. As mentioned in [31], it was unknown yet whether such a secure MSR code is still optimal when $l_2 \geq 1$. Since then, characterization of the secrecy capacity for MSR codes is considered to be open under $(l_1, l_2 > 0)$ -eavesdropper model.

Recently, the authors in [32] and [33] employ the technique of linear subspace intersection and then derive new upper bounds on secrecy capacity for linear MSR codes. Zigzag code [23] and its variant [21] are shown to achieve these bounds through pre-coding of maximum rank distance (MRD) code [36,37]. The bound given in [33] auxilarily implies that the product-matrix-based secure MSR code proposed in [31] is also optimal for $l_2 = 1$. Regarding the bound given in [32], it is actually an extension of the one in [33], since the bound in [32] matches to that in [33] when $l_2 \leq 2$.

In another parallel research area, towards two separate eavesdropper models with $(l_1, l_2 = 0)$ and $(l_1 = 0, l_2)$, the authors in [34] study the secure storage-vs-repair-bandwidth tradeoff, where they respectively derive new outer bounds on secrecy capacity for a general parameter set and some specific parameter sets. Therein, they show that in the presence of $(l_1 = 0, l_2)$ -eavesdropper, these new bounds strictly improve upon the existing cutset-based bounds presented in [30] and the MBR point is the only efficient point that can achieve these specific-parameter-based bounds. Under the above background of (l_1, l_2) -eavesdropper model, our focus herein is dedicated to studying the secrecy capacity solely at the MSR point².

Contributions: In this work, we first carefully review the method of determining regenerating codes considered in [7] and the information-theoretic technique used in [9]. Therein, we find that the α symbols stored in any node or the β symbols contained in any single set of repair data for the *optimal* regenerating codes are in fact mutually independent and uniformly distributed inside themselves. It not only indicates that entropy of any symbol involved reaches the maximal value 1, but also signifies that entropy of the α symbols in any node and entropy of the β symbols in any single repair data all attain the maximal-integer-value α and β respectively. Thereafter, we recognize that the concepts of uniform distribution and independence between symbols in information theory [1] exactly correspond to those of permutation polynomial and orthogonal system in finite fields [2] respectively. Using these two theories in finite fields,

² It is shown in [34] that for certain parameters, secure codes operating at the MBR points actually have better “storage” (i.e., the maximal file size that can be securely stored, or just termed secrecy capacity) rate than codes operating at the MSR points. In this sense, it appears that secure MSR codes lose the feature of optimal storage, while the original notion of MSR codes under the non-secure setting shall be optimal in storage rate as displayed in [7]. Throughout this paper, we still use the term MSR points (or codes) to only signify the fact that α and β satisfy the relationship $\alpha = (d - k + 1)\beta$ like the MBR points termed in [34] that require $\alpha = d\beta$. Essentially, each node in the secure MSR codes still stores α_{MSR} symbols and transmits β_{MSR} symbols for repairing failed nodes, which just need to replace with some randomness.

we demonstrate that the joint entropy of symbols included in multiple sets of repair data in the nonlinear context may be a non-integer value while it must be an integer in the linear context, which will be used to investigate the secrecy capacity of linear MSR codes.

Then, we turn to study the inherent features of MSR codes from the information-theoretic perspective, where the data stored in storage nodes and transferred by helper nodes during repair are considered as random variables. Based on the basic reconstruction and regeneration properties of MSR codes with $\{n = d + 1, k, d, \alpha, \beta\}$, we derive two useful properties: (i) the repair data sent from disjoint sets of nodes to a failed node are mutually independent, and (ii) given the contents of a node and the repair data from any $k - 1$ nodes, the repair data from the remaining $d - k + 1$ nodes are deterministic. Combining the two new properties with a universal upper bound on secrecy capacity for any *optimal* regenerating code with $\{n = d + 1, k, d, \alpha, \beta\}$, we derive a simple and generally applicable upper bound on secrecy capacity for any MSR code with $\{n = d + 1, k, d, \alpha, \beta\}$. As for the MSR codes with $\{n > d + 1, k, d, \alpha, \beta\}$, we introduce a new concept of “*stable*” MSR codes, which require that repair data transmitted from any node i to any failed node j is independent of the choice of the set of helper nodes including the same node i . Therein, we show this stable property is the equivalent condition of secrecy capacity between any MSR code with $\{n > d + 1, k, d, \alpha, \beta\}$ and its truncated one with $\{n = d + 1, k, d, \alpha, \beta\}$. It should be noted that the product-matrix-based MSR code given in [15] is a *stable* MSR code.

Finally, we converge back to the linear MSR codes with parameter set $\{n = d + 1, k, d, \alpha, \beta\}$, where those aforementioned upper bounds on secrecy capacity actually can always be achieved through the pre-coding of maximum rank distance (MRD) code [36,37] as applied in [33,35]. Based on the fact that joint entropy of multiple sets of repair data is an integer, we fully characterize the secrecy capacity of linear MSR codes in the category where $1 \leq \beta < \frac{d-k+1}{l_2-1}$. A consequence of this result when $\beta = 1$ naturally establishes the optimality of product-matrix-based secure MSR codes whenever $l_1 + l_2 \leq k - 1$ and $l_2 \leq \min\{k - 1, d - k + 1\}$, which completely resolves the question raised in [31]. Note that product-matrix-based MSR code given in [15] is a scalar MSR code, i.e., it is built on $\beta = 1$. In the other category where $\beta \geq \frac{d-k+1}{l_2-1}$, we give new upper bounds on secrecy capacity, which are in fact improved generalization of the results given in [32,33]. Thereafter, we find that all the aforementioned results also apply to systematic MSR codes with only repair data of systematic nodes eavesdropped. By putting all together, we eventually present a comprehensive and explicit result on secrecy capacity for linear MSR codes, which closely depends on the value of β . This final outcome also predicts certain results on secrecy capacity for some unexplored linear MSR codes. As an illustration and comparison, Table 1 summarizes the study progresses on secrecy capacity for linear MSR codes, wherein it should be noted that the bound in [34] cannot be reached for MSR codes.

Table 1. Secrecy Capacity of Linear MSR Codes under (l_1, l_2) -Eavesdropper Model

Citation	Corresponding Results
Pawar et al [30]	$B^{(s)} \leq (k - l_1 - l_2)\alpha$, optimal only when $l_2 = 0$
Tandon et al [34]	$B^{(s)} \leq (k - l_2)(1 - \frac{1}{d})\alpha$, for $n = d + 1$, $l_1 = 0$ and $1 \leq l_2 < k$
Shah et al [31]	$B^{(s)} = (k - l_1 - l_2)(\alpha - l_2\beta)$, for product-matrix-based MSR codes
Rawat et al [33]	$B^{(s)} \leq (k - l_1 - l_2)(\alpha - \theta(\beta, l_2))$, where $\theta(\beta, l_2) = \begin{cases} \beta, & \text{for } l_2 = 1 \\ 2\beta - \frac{\beta}{d+1-k}, & \text{for } l_2 = 2 \end{cases}$
Goparaju et al [32]	$B^{(s)} \leq (k - l_1 - l_2)(1 - \frac{1}{n-k})^{l_2}\alpha$, where $n = d + 1$
This paper	$B^{(s)} = \underbrace{(k - l_1 - l_2)(\alpha - \pi(\beta, l_2))}_{\text{wherein}}$, wherein $\pi(\beta, l_2): \begin{cases} = l_2\beta, & \text{if } l_2 \leq t, \beta < \frac{d-k+1}{t-1}; \\ \geq t\beta + \beta(d - k - t + 1)[1 - (\frac{d-k}{d-k+1})^e], & \text{if } l_2 = t + e, \frac{d-k+1}{t} \leq \beta < \frac{d-k+1}{t-1}, \end{cases}$ where $1 \leq t \leq d - k + 1$ and $e \geq 1$. This also can be referenced from our formula (71)

Organization: Section 2 gives preliminaries consisting of some basic definitions in information theory, notation used in this paper, some results from theory of finite field and a universal upper bound under the (l_1, l_2) -eavesdropper model. Section 3 presents some new results for general MSR codes mainly including some general properties, some generally applicable upper bounds on secrecy capacity and the new concept of *stable* MSR codes. Section 4 exhibits the comprehensive and explicit result on secrecy capacity for linear MSR codes. Section 5 concludes this paper.

2 PRELIMINARIES

In this section, some basic concepts related to information theory are quoted, which will be used in high frequency later. Then, we describe the system model of MSR codes from the information-theoretic perspective. Subsequently, we introduce the theory on permutation polynomial in finite fields, which can be regarded as a new way to understand the construction of *optimal* regenerating codes. At last, we present a universal upper bound on secrecy capacity under the (l_1, l_2) -eavesdropper model.

2.1 Information Entropy

Definition 1. [1](Entropy of A Random Variable X): The entropy of a discrete random variable X with probability distribution $p_X(x)$ is

$$H(X) = - \sum_x p(x) \log p(x). \quad (4)$$

The entropy measures the expected uncertainty in X . It must be that $H(X) \geq 0$, meaning entropy is always non-negative and $H(X) = 0$ iff X is deterministic. In addition, when X is uniformly distributed (i.e., $p(x) = \frac{1}{q}$ where q is the total number of the events of X), $H(X)$ achieves the maximum value $\log q$. Normally, the base of logarithm can be specified to q . In this case, it must be that $H(X) \leq 1$ and $H(X) = 1$ iff X is uniformly distributed.

Definition 2. [1](Joint Entropy and Conditional Entropy): Joint entropy between two random variables X and Y , and conditional entropy of Y given a random variable X are respectively

$$\begin{cases} H(X, Y) = -E_{p(x,y)}[\log p(X, Y)] = - \sum_x \sum_y p(x, y) \log p(x, y) \\ H(Y|X) = -E_{p(x,y)}[\log p(Y|X)] = - \sum_x p(x) H(Y|X = x) \end{cases} \quad (5)$$

Besides, joint and conditional entropy provide a natural calculus: $H(X, Y) = H(X) + H(Y|X)$.

Definition 3. [1](Mutual Information and Conditional Mutual Information): The mutual information between X and Y , and the conditional mutual information between X and Y given another random variable Z are respectively given by:

$$\begin{cases} I(X; Y) = H(X) - H(X|Y) \\ I(X; Y|Z) = H(X|Z) - H(X|Y, Z) \end{cases} \quad (6)$$

Definition 4. [1](Chain Rules): Chain rules for entropy and mutual information are:

$$\begin{cases} H(X_1, \dots, X_n) = \sum_{i=1}^n H(X_i | X_{i-1}, \dots, X_1) \\ I(X_1, \dots, X_n; Y) = \sum_{i=1}^n I(X_i; Y | X_{i-1}, X_{i-2}, \dots, X_1) \end{cases} \quad (7)$$

Lemma 1. Based on these definitions of information entropy, we naturally have

$$I(X; Y|Z) = I(Y; X|Z) \leq \min\{H(X), H(Y)\}. \quad (8)$$

2.2 Notation

We follow the information-theoretic approach introduced in [9] and accordingly treat all data symbols including data stored at the storage nodes and those transferred by helper nodes during the repair operations as random variables.

Note 1 Throughout the paper, we mainly consider the situation of MSR code with parameter set $\{n = d + 1, k, d, \alpha, \beta\}$, because any upper bound on the data file that can be securely stored for any secure MSR code with $\{n = d + 1, k, d, \alpha, \beta\}$ also holds for the corresponding secure MSR code with $\{n > d + 1, k, d, \alpha, \beta\}$. In Section 3.3, we will establish the equivalent condition of secrecy capacity between any MSR code with $\{n > d + 1, k, d, \alpha, \beta\}$ and its truncated one with $\{n = d + 1, k, d, \alpha, \beta\}$.

We represent nodes using indices 1 to n and denote the sequence of nodes $[i, i + 1, \dots, j]$ by $[i, j]$, where $i < j$. We use symbols for a set $\{\dots\}$ and a sequence $[\dots]$ interchangeably. For any regenerating code with parameter set $\{n = d + 1, k, d, \alpha, \beta\}$, we let

■ (1). $W_i, i \in [1, d + 1]$ denote the random variable corresponding to the content of node i . As proved in [9], it must be that $H(W_i) = \alpha$ for any *optimal* regenerating code including MSR codes.

■ (2). $\{W_A, A \subseteq [1, d + 1]\}$ denote the set of random variables corresponding to the nodes in the subset A . Throughout the paper, subscripts of W can represent either a node index or a set of nodes which will be clear from the context.

■ (3). $S_i^j, \{i, j\} \in [1, d + 1], i \neq j$ denote the random variable corresponding to the data symbols sent by the helper node i to the replacement of the failed node j . It must be that $H(S_i^j) = \beta$ for any *optimal* regenerating code including MSR codes, following from [9].

■ (4). S_A^B denote the set $\{S_i^j | i \in A, j \in B, i \neq j, A \subseteq [1, d + 1], B \subseteq [1, d + 1]\}$, and particularly S^B substitutes for $S_{[1, d + 1]}^B$.

According to the above notation, reconstruction as well as regeneration property of any regenerating code can be expressed as

$$\begin{cases} H(W_{i_1}, W_{i_2}, \dots, W_{i_k}) = k\alpha, & i_j \in [1, d + 1], j \in \{1, \dots, k\} \\ H(W_i | S_{\{[1, d + 1] \setminus i\}}^i) = 0, & i \in [1, d + 1] \end{cases} \quad (9)$$

2.3 Permutation Polynomials

As shown in [7], α represents the number of symbols stored in each node and β corresponds to the number of symbols downloaded from a surviving node to repair a failed node. Note that the entropy of each symbol cannot be greater than 1 and may not be an integer. Thus, it can only be that $H(W_i) \leq \alpha$ and $H(S_i^j) \leq \beta$. Subsequently, under the context of *optimal* regenerating codes, Shah et al in [9] employ information theory to derive that $H(W_i) = \alpha$ and $H(S_i^j) = \beta$, which implies that each symbol contained in each node and repair data actually reaches the maximum entropy 1, i.e., each symbol is uniformly distributed inside itself. Besides, it also means that the symbols included in the same node i and same repair data S_i^j are mutually independent respectively. Although each symbol included in any repair data S_i^j has the uniform distribution and S_i^j also has the maximal entropy β , the joint entropy $H(S_i^{j_1}, S_i^{j_2})$ may not be an integer where $j_1 \neq j_2$ as illustrated in the following.

We let $(y_1^i, y_2^i, \dots, y_\alpha^i)$ denote the α symbols stored in node i , where $H(y_l^i) = 1$ for any $l \in [1, \alpha]$. In addition, we let $(z_1^{(i, j_1)}, z_2^{(i, j_1)}, \dots, z_\beta^{(i, j_1)})$ and $(z_1^{(i, j_2)}, z_2^{(i, j_2)}, \dots, z_\beta^{(i, j_2)})$ be the β symbols contained in the repair data $S_i^{j_1}$ and $S_i^{j_2}$ respectively, where $H(z_l^{(i, j_1)}) = H(z_l^{(i, j_2)}) = 1$ for $l \in [1, \beta]$. Now consider the joint entropy

$$H(S_i^{j_1}, S_i^{j_2}) = H(z_1^{(i, j_1)}, z_2^{(i, j_1)}, \dots, z_\beta^{(i, j_1)}, z_1^{(i, j_2)}, z_2^{(i, j_2)}, \dots, z_\beta^{(i, j_2)}). \quad (10)$$

In a finite field \mathbb{F}_q , any mapping $\tau : \mathbb{F}_q^n \rightarrow \mathbb{F}_q$ can be represented by a polynomial over \mathbb{F}_q of degree $< q$ in each “indeterminate” through Lagrange Interpolation [2]. Since all the symbols contained in

node i are uniformly distributed inside themselves and mutually independent, they can be regarded as “indeterminates”. So, we let

$$\begin{cases} z_1^{(i,j_1)} = \dot{f}_1(y_1^i, y_2^i, \dots, y_\alpha^i) \\ z_2^{(i,j_1)} = \dot{f}_2(y_1^i, y_2^i, \dots, y_\alpha^i) \\ \vdots \\ z_\beta^{(i,j_1)} = \dot{f}_\beta(y_1^i, y_2^i, \dots, y_\alpha^i) \end{cases} \quad (11)$$

and

$$\begin{cases} z_1^{(i,j_2)} = \ddot{f}_1(y_1^i, y_2^i, \dots, y_\alpha^i) \\ z_2^{(i,j_2)} = \ddot{f}_2(y_1^i, y_2^i, \dots, y_\alpha^i) \\ \vdots \\ z_\beta^{(i,j_2)} = \ddot{f}_\beta(y_1^i, y_2^i, \dots, y_\alpha^i), \end{cases} \quad (12)$$

where $(\dot{f}_1, \dot{f}_2, \dots, \dot{f}_\beta)$ and $(\ddot{f}_1, \ddot{f}_2, \dots, \ddot{f}_\beta)$ represent the polynomials induced by the symbols contained in repair data $S_i^{j_1}$ and $S_i^{j_2}$ respectively. In [2], there are two special concepts introduced as follows.

Definition 5. [2](Permutation Polynomial): A polynomial $f \in \mathbb{F}_q[x_1, \dots, x_n]$ is called a permutation polynomial in n indeterminates over \mathbb{F}_q if the equation $f(x_1, \dots, x_n) = a$ has q^{n-1} solutions in \mathbb{F}_q^n for each $a \in \mathbb{F}_q$.

According to Definition 5, we know that each value $a \in \mathbb{F}_q$ will be taken in the same probability ($\frac{q^{n-1}}{q^n} = \frac{1}{q}$) by a permutation polynomial. From this point, permutation polynomial exactly corresponds to uniform distribution in information theory (Definition 1). Due to that $H(z_l^{(i,j_1)}) = H(z_l^{(i,j_2)}) = 1$ for any $l \in [1, \beta]$, we know that $(\dot{f}_1, \dot{f}_2, \dots, \dot{f}_\beta, \ddot{f}_1, \ddot{f}_2, \dots, \ddot{f}_\beta)$ all are permutation polynomials. Here, it should be noted that permutation polynomials are not necessarily linear polynomials in finite fields while linear polynomials apparently are permutation polynomials.

Definition 6. [2](Orthogonal System): A system of polynomials $f_1, \dots, f_m \in \mathbb{F}_q[x_1, \dots, x_n]$ where $1 \leq m \leq n$ is said to be orthogonal in \mathbb{F}_q , if the system of equations

$$\begin{cases} f_1(x_1, \dots, x_n) = a_1 \\ \vdots \\ f_m(x_1, \dots, x_n) = a_m \end{cases} \quad (13)$$

has q^{n-m} solutions in \mathbb{F}_q^n for each $(a_1, \dots, a_m) \in \mathbb{F}_q^m$.

According to Definition 6 and Definition 2, we know that $(\dot{f}_1, \dot{f}_2, \dots, \dot{f}_\beta)$ and $(\ddot{f}_1, \ddot{f}_2, \dots, \ddot{f}_\beta)$ respectively constitute two orthogonal systems, since $H(S_i^{j_1}) = H(S_i^{j_2}) = \beta$. Similarly, it follows that $H(S_i^{j_1}, S_i^{j_2}) = 2\beta$ if and only if the 2β polynomials $(\dot{f}_1, \dot{f}_2, \dots, \dot{f}_\beta, \ddot{f}_1, \ddot{f}_2, \dots, \ddot{f}_\beta)$ can form an orthogonal system.

However, if there exist two different polynomials \dot{f}_{l_1} and \ddot{f}_{l_2} for some $l_1, l_2 \in [1, \beta]$ that cannot form an orthogonal system, the joint entropy of the corresponding symbols $H(z_{l_1}^{(i,j_1)}, z_{l_2}^{(i,j_2)})$ will be a non-integer, which may result in that all the symbols of repair data $S_i^{\{j_1, j_2\}}$ also have the non-integer joint entropy. Note that multiple permutation polynomials may not form an orthogonal system while each polynomial in an orthogonal system must be a permutation polynomial.

Nevertheless, in the linear context, the joint entropy of the symbols contained in S_i^A must be an integer, where $i \notin A$ and A is any subset of $[1, d+1]$.

Lemma 2. *In the scenario of linear optimal regenerating codes, the symbols contained in S_i^A must have integer-value entropy, where $i \in [1, d+1]$, $A \subset [1, d+1]$ and $i \notin A$.*

Proof. Assume all the $m = |A| \cdot \beta$ symbols in S_i^A can be represented as

$$\{f_1(x_1, x_2, \dots, x_\alpha), f_2(x_1, x_2, \dots, x_\alpha), \dots, f_m(x_1, x_2, \dots, x_\alpha)\}, \quad (14)$$

where $(x_1, x_2, \dots, x_\alpha)$ are the α symbols stored in node i and f_l denotes the linear polynomial for $l \in [1, m]$. Then, we let

$$\begin{cases} f_1(x_1, x_2, \dots, x_\alpha) = a_{11}x_1 + a_{12}x_2 + \dots + a_{1\alpha}x_\alpha \\ f_2(x_1, x_2, \dots, x_\alpha) = a_{21}x_1 + a_{22}x_2 + \dots + a_{2\alpha}x_\alpha \\ \vdots \\ f_m(x_1, x_2, \dots, x_\alpha) = a_{m1}x_1 + a_{m2}x_2 + \dots + a_{m\alpha}x_\alpha, \end{cases} \quad (15)$$

where all the coefficients are drawn from \mathbb{F}_q . Equation (15) can be alternatively expressed as

$$(f_1, f_2, \dots, f_m)^T = C \cdot (x_1, x_2, \dots, x_\alpha)^T, \quad (16)$$

where T indicates the transpose operation and C denotes the generator matrix.

Since (f_1, f_2, \dots, f_m) are linear combinations of a set of uniformly distributed random variables, then they all are uniformly distributed and they are either mutually independent, or some of them are determined by the remaining of them. In fact, the value of $H(S_i^A)$ is equal to the rank of C , which we denote by $r(C)$.

1. When $m \leq \alpha$ and $r(C) = m$, the row vectors of C are linearly independent. Then, for each vector value $(b_1, b_2, \dots, b_m) \in \mathbb{F}_q^m$, equation (15) has $q^{\alpha-m}$ solutions in \mathbb{F}_q^α . Thus, each vector value (b_1, b_2, \dots, b_m) will occur in equally probability $\frac{q^{\alpha-m}}{q^\alpha} = \frac{1}{q^m}$. So, the polynomials (15) form an orthogonal system. In this case, we can calculate that $H(S_i^A) = m = r(C)$ according to Definition 2.

2. When $r(C) < m$, the row vectors of C are not linearly independent, which implies $r(C)$ chosen linearly independent polynomials f_l will determine the values of the remaining $m - r(C)$ polynomials. Although the whole polynomials (15) cannot form the orthogonal system, the $r(C)$ linearly independent polynomials still forms an orthogonal system. Similar to the above case, entropy of these $r(C)$ linearly independent polynomials is equal to $r(C)$. Thereby, we have

$$H(S_i^A) = H(f_1, f_2, \dots, f_m) = H(f_{l_1}, f_{l_2}, \dots, f_{l_{r(C)}}) = r(C), \quad (17)$$

where $\{f_{l_1}, f_{l_2}, \dots, f_{l_{r(C)}}\}$ are the $r(C)$ chosen linearly independent polynomials.

Hence, both cases indicate that $H(S_i^A) = r(C)$, while $r(C)$ must be an integer since it represents the rank of C .

Remark 1 *In this lemma, theories of permutation polynomial and orthogonal system are used to demonstrate that the joint entropy of symbols included in multiple sets of repair data in the nonlinear context may be a non-integer value while their joint entropy has to be an integer in the linear context, which is important for the later discussion on secrecy capacity of linear MSR codes.*

Additionally, it is of independent interest that these two theories in finite fields are also applicable to the nonlinear context, because they may be utilized to explore the case of constructing nonlinear optimal regenerating codes. That is beyond the scope of this paper though. In this paper, we mainly study the secrecy capacity of linear MSR codes, while some new insights on general MSR codes are also present.

2.4 A Universal Upper Bound

► **Eavesdropper Model:** Let E be a set of l_1 nodes which the eavesdropper has access to, and F be another disjoint set of l_2 nodes whose repair data can be observed by the eavesdropper. In other words, the eavesdropper is assumed to have the knowledge of $\{W_E, S^F\}$. Furthermore, we assume $l_1 + l_2 < k$, otherwise the eavesdropper can retrieve all the data message. Due to this eavesdropper model, we set G to be another subset $G \subseteq \{[1, d+1] \setminus (E \cup F)\}$ of size $(k - l_1 - l_2)$. Based on this model, a universal upper bound on the secrecy capacity of any *optimal* regenerating code is given as follows.

Lemma 3. *For any secure optimal regenerating code with $\{n = d + 1, k, d, \alpha, \beta\}$, we have*

$$\begin{cases} B^{(s)} \\ \leq H(W_E, W_F, W_G | W_E, S^F) \\ = H(W_G | W_E, W_F) - H(S^F | W_E, W_F) \\ = \sum_{i=l_1+l_2+1}^k \min\{\alpha, (d-i+1)\beta\} - H(S^F | W_E, W_F) \end{cases} \quad (18)$$

Proof. First, in secure regenerating codes [31,32,33], the random variables associated with the message can be viewed as the tuple (D, R) , where D corresponds to the actual data file and R corresponds to the randomness added. The secure file size is $B^{(s)} = H(D)$ and the secrecy condition requires that $I(D; W_E, S^F) = 0$. Thus, it must be that

$$\begin{cases} H(D) \\ = H(D) - I(D; W_E, S^F) \\ = H(D | W_E, S^F) \\ \leq H(D, R | W_E, S^F) \\ = H(W_E, W_F, W_G | W_E, S^F), \end{cases} \quad (19)$$

where the equation in the last step follows from the reconstruction property.

Second, we have

$$\begin{cases} H(W_G | W_E, W_F) - H(W_E, W_F, W_G | W_E, S^F) \\ = H(W_G | W_E, W_F) - H(W_E, W_F, W_G | W_E, W_F, S^F) \\ = H(W_G | W_E, W_F) - H(W_G | W_E, W_F, S^F) \\ = I(W_G; S^F | W_E, W_F) \\ = H(S^F | W_E, W_F) - H(S^F | W_E, W_F, W_G) \\ = H(S^F | W_E, W_F), \end{cases} \quad (20)$$

where the regeneration property leads to $H(S^F) = H(S^F, W_F)$ that is used in the first step.

At last, for the *optimal* regenerating codes, it follows from the proof of property 1 in [9] that

$$H(W_G | W_E, W_F) = \sum_{i=l_1+l_2+1}^k \min\{\alpha, (d-i+1)\beta\}. \quad (21)$$

Remark 2 *In the context of linear regenerating codes, MRD (Maximum Rank Distance) codes [36] (e.g. Gabidulin code [37]) can be used to pre-code the original data of size $\{B = k\alpha\}$, which is required to consist of $\{B - H(W_E, S^F)\}$ -sized actual data file D and $H(W_E, S^F)$ -sized random data R . It should be noted that $H(W_E, S^F)$ is also an integer as derived in Lemma 2, because $\{W_E, S^F\}$ are obtained by the*

linear combinations of the original data message of size B . As shown in [33,35], this kind of secure code construction always can meet the secrecy condition that $I(D; W_E, S^F) = 0$. It exactly means the maximal file size that can be securely stored is

$$B^{(s)} = B - H(W_E, S^F) = H(W_E, W_F, W_G | W_E, S^F). \quad (22)$$

In the MSR scenario, it is obvious that $H(W_G | W_E, W_F) = H(W_G) = (k - l_1 - l_2)\alpha$, following from property 2 in [9]. Thus, we only need to concentrate on the term $H(S^F | W_E, W_F)$ in this paper.

3 DATA SECURITY FOR GENERAL MSR CODES

In this section, we give some general properties of MSR codes and a simple expression of upper bound on secrecy capacity, which will be leveraged throughout this paper. Afterwards, *stable* MSR code as a new concept is introduced, where the stable property will be shown to be closely linked with secrecy capacity.

3.1 Properties of General MSR Codes

Here, we proceed to provide some new properties of general MSR codes (including the nonlinear context), which actually stem from the reconstruction and regeneration properties of MSR codes. With these properties, we can further simplify the formulation $H(S^F | W_E, W_F)$ mentioned above.

Lemma 4. *In the scenario of MSR codes with parameter set $\{n = d + 1, k, d, \alpha, \beta\}$, for any node i with efficient repair, consider two arbitrary subsets A' and B' such that $\{|A'| = k - 1, |B'| = d - k + 1, A' \cap B' = \emptyset, A' \cup B' = [1, d + 1] \setminus i\}$, it must be that*

$$\begin{cases} H(S_{A' \cup B'}^i) = d\beta \\ H(S_{B'}^i | W_i, S_{A'}^i) = 0. \end{cases} \quad (23)$$

Proof. Without loss of generality, we assume $i = 1$. The proof is given in two steps as follows.

1. According to the Property 2 in [9], it is trivial that $I(W_1; W_{A'}) = 0$ in the MSR scenario, which leads to $H(W_1 | S_{A'}^1) = \alpha$.

2. We set $B' = (b_1, b_2, \dots, b_{d-k+1})$. Due to the repair property, it must be that $H(W_1 | S_{A'}^1, S_{B'}^1) = 0$. Next, some key inequalities are present from Lemma 1:

$$\left\{ \begin{array}{l} H(W_1 | S_{A'}^1) - H(W_1 | S_{A'}^1, S_{b_1}^1) \\ = I(W_1; S_{b_1}^1 | S_{A'}^1) \\ = H(S_{b_1}^1 | S_{A'}^1) - H(S_{b_1}^1 | W_1, S_{A'}^1) \\ \leq \beta; \\ H(W_1 | S_{A'}^1, S_{b_1}^1) - H(W_1 | S_{A'}^1, S_{b_1}^1, S_{b_2}^1) \\ = I(W_1; S_{b_2}^1 | S_{A'}^1, S_{b_1}^1) \\ = H(S_{b_2}^1 | S_{A'}^1, S_{b_1}^1) - H(S_{b_2}^1 | W_1, S_{A'}^1, S_{b_1}^1) \\ \leq \beta; \\ \vdots \\ H(W_1 | S_{A'}^1, S_{b_1}^1, S_{b_2}^1, \dots, S_{b_{d-k}}^1) - H(W_1 | S_{A'}^1, S_{B'}^1) \\ = I(W_1; S_{b_{d-k+1}}^1 | S_{A'}^1, S_{\{B' \setminus b_{d-k+1}\}}^1) \\ = H(S_{b_{d-k+1}}^1 | S_{A'}^1, S_{\{B' \setminus b_{d-k+1}\}}^1) - H(S_{b_{d-k+1}}^1 | W_1, S_{A'}^1, S_{\{B' \setminus b_{d-k+1}\}}^1) \\ \leq \beta. \end{array} \right. \quad (24)$$

By summing up the left side of the inequalities, we derive

$$\alpha = H(W_1|S_{A'}^1) - H(W_1|S_{A'}^1, S_{B'}^1) \leq (d - k + 1)\beta. \quad (25)$$

Because $\alpha = (d - k + 1)\beta$, it is mandatory that all the inequalities (24) actually are equations. Thus, for any $j \in [1, d - k + 1]$, we have

$$\begin{cases} H(S_{b_j}^1|S_{A'}^1, S_{\{b_1, \dots, b_{j-1}\}}^1) = \beta \\ H(S_{b_j}^1|W_1, S_{A'}^1, S_{\{b_1, \dots, b_{j-1}\}}^1) = 0, \end{cases} \quad (26)$$

from which we can derive

$$\begin{cases} H(S_{A' \cup B'}^1) \\ = H(S_{A'}^1) + H(S_{B'}^1|S_{A'}^1) \\ = (k - 1)\beta + \sum_{j=1}^{d-k+1} H(S_{b_j}^1|S_{A'}^1, S_{\{b_1, \dots, b_{j-1}\}}^1) \\ = (k - 1)\beta + (d - k + 1)\beta \\ = d\beta \end{cases} \quad (27)$$

and

$$\begin{cases} H(S_{B'}^1|W_1, S_{A'}^1) \\ = \sum_{j=1}^{d-k+1} H(S_{b_j}^1|W_1, S_{A'}^1, S_{\{b_1, \dots, b_{j-1}\}}^1) \\ = 0 \end{cases} \quad (28)$$

Remark 3 This lemma exhibits the special properties of MSR codes. $H(S_{A' \cup B'}^i) = d\beta$ means any repair data from disjoint sets of nodes upon failure of node i are mutually independent. $H(S_{B'}^i|W_i, S_{A'}^i) = 0$ implies that given the contents of node i and the repair data from any $k - 1$ nodes, the repair data from the remaining $d - k + 1$ nodes are deterministic.

Lemma 5. In the MSR scenario with $\{n = d + 1, k, d, \alpha, \beta\}$, we have

$$H(S^F|W_E, W_F) = H(S_G^F), \quad (29)$$

where E, F and G are pairwise disjoint sets as defined in Section 2.4 and $|E \cup F \cup G| = k$. Furthermore, when $E = \emptyset$, we still have $H(S^F|W_F) = H(S_G^F)$, where $|F \cup G| = k$.

Proof. Assume all the $d + 1$ nodes are comprised of E, F, G and T , where $|E \cup F \cup G| = k$ and $|T| = d - k + 1$. Thereby, we have

$$\begin{cases} H(S^F|W_{\{E, F\}}) \\ = H(S_{\{E, F, G, T\}}^F|W_{\{E, F\}}) \\ = H(S_{\{G, T\}}^F|W_{\{E, F\}}) \\ = H(S_G^F|W_{\{E, F\}}) + H(S_T^F|W_{\{E, F\}}, S_G^F). \end{cases} \quad (30)$$

Then, due to the condition $|(E, F, G) \setminus i| = k - 1$, Lemma 4 leads to that for any $i \in F$,

$$\begin{cases} H(S_T^i|W_{\{E, F\}}, S_G^i) \\ \leq H(S_T^i|W_{\{E, F\}}, S_G^i) \\ = H(S_T^i|W_i, W_{\{(E, F) \setminus i\}}, S_G^i) \\ \leq H(S_T^i|W_i, S_{\{(E, F) \setminus i\}}^i, S_G^i) \\ = H(S_T^i|W_i, S_{\{(E, F, G) \setminus i\}}^i) \\ = 0, \end{cases} \quad (31)$$

from which we derive $H(S_T^F | W_{\{E,F\}}, S_G^F) = 0$.

Still by the Property 2 in [9], it has to be that $H(S_G^F | W_E, W_F) = H(S_G^F)$, since $|E \cup F \cup G| = k$. Hence, we obtain the proof. In addition, the above deduction is obviously applicable to the situation when $E = \emptyset$.

Remark 4 Combining this lemma with Lemma 3, we have

$$B^{(s)} \leq (k - l_1 - l_2)\alpha - H(S_G^F). \quad (32)$$

In particular, for the linear MSR codes, this upper bound can always be achieved (as in Remark 2). Moreover, the equation $H(S^F | W_F) = H(S_G^F)$ promotes the next result.

Lemma 6. In the MSR scenario with $\{n = d + 1, k, d, \alpha, \beta\}$, for any subset F such that $|F| \leq k - 1$, and arbitrary different i_1, i_2 where $i_1, i_2 \notin F$, we have $H(S_{i_1}^F) = H(S_{i_2}^F)$.

Proof. According to Lemma 5, we obtain

$$\begin{cases} H(S^F) \\ = H(S^F, W_F) \\ = H(W_F) + H(S^F | W_F) \\ = H(W_F) + H(S_{G'}^F | W_F) \\ = H(W_F) + H(S_{G'}^F), \end{cases} \quad (33)$$

where G' is any subset of $[1, d + 1]$ such that $|G'| + |F| = k$ and $G' \cap F = \emptyset$. Based on the condition $|F| \leq k - 1$, then it has to be that $|G'| \geq 1$.

1. When $|G'| = 1$, for any two different g_1 and g_2 where $g_1, g_2 \in \{[1, d + 1] \setminus F\}$,

$$H(S^F) = H(W_F) + H(S_{g_1}^F) = H(W_F) + H(S_{g_2}^F), \quad (34)$$

which indicates $H(S_{g_1}^F) = H(S_{g_2}^F)$.

2. When $|G'| \geq 2$, we set $G' = \{g', G_1\}$ and $G'' = \{g'', G_1\}$ such that $\{g' \neq g'', |G'| = |G''| = k - |F|, G' \cap F = G'' \cap F = \emptyset\}$, where G'' plays the same role of G' in the following statement. Similarly, we derive

$$\begin{cases} H(S^F) \\ = H(W_F) + H(S_{G'}^F) \\ = H(W_F) + H(S_{g'}^F) + H(S_{G_1}^F); \\ H(S^F) \\ = H(W_F) + H(S_{G''}^F) \\ = H(W_F) + H(S_{g''}^F) + H(S_{G_1}^F), \end{cases} \quad (35)$$

which implies $H(S_{g'}^F) = H(S_{g''}^F)$.

Because the choice of (g_1, g_2) and (g', g'') are arbitrary, then for arbitrary different i_1, i_2 where $i_1, i_2 \notin F$, we have $H(S_{i_1}^F) = H(S_{i_2}^F)$.

Remark 5 Lemma 6 shows that the entropy of repair data from any two nodes assisting in repairing the same subsets of nodes are identical. Combining this lemma with Remark 4, we further obtain the following result on upper bound.

3.2 A Simple Expression of Upper Bound

Incorporating Lemma 3, Lemma 5 and Lemma 6, we consequently derive a simple and generally applicable result on secrecy capacity as follows.

Theorem 1. *In the scenario of MSR codes with parameter set $\{n = d + 1, k, d, \alpha, \beta\}$, we have*

$$B^{(s)} \leq (k - l_1 - l_2)(\alpha - H(S_g^F)), \quad (36)$$

where $g \in G$, $|G| = k - l_1 - l_2$, and $|F| = l_2$.

Remark 6 *This can be viewed as a simple and generally applicable upper bound of $B^{(s)}$, since we only need to calculate or estimate the joint entropy of repair data transmitted from any single node $H(S_g^F)$. Still by Remark 2, this upper bound can be reached in the scenario of linear MSR codes.*

3.3 Stable MSR Codes

Given an MSR code with $\{n = d + 1, k, d, \alpha, \beta\}$, since our focus is the exact repair, the random variables W_j are invariant with time, i.e., they remain constant irrespective of the sequence of failures and repairs that occur in the storage system. Once construction of such an MSR code with $\{n = d + 1\}$ is present, content of S_i^j sent from a node i to repair another node j also keeps invariant. However, for the MSR code with $\{n > d + 1, k, d, \alpha, \beta\}$, the repair data S_i^j technically need not keep constant and may vary with different sets of helper nodes including the same node i , only needing to satisfy that per node storage W_j stays unchanged. For instance, when node j is failed, node i is assigned to assist in repairing node j . Thus, there totally exists $\binom{d-1}{n-2}$ possible sets of helper nodes including node i .

Assume repair data of node j is captured by the eavesdropper³. If content of repair data S_i^j is not independent of the choice of the set of helper nodes and varies with them, after multiple repair epochs with different sets of helper nodes including node i , different information regarding repair data S_i^j will be exposed to the eavesdropper. Thus, the eavesdropper is supposed to observe more information regarding repair data of node j , when compared to the case of invariant content of repair data. In the following, we will use an example to illustrate this security issue.

Example 1 *Assume $E = \emptyset$ and $F = \{1\}$, i.e., only the repair data of node 1 is eavesdropped. Consider two truncated MSR codes \mathbb{M} and \mathbb{M}' comprised of nodes set $[1, d + 1]$ and $[1, 3, \dots, d + 2]$ respectively from an MSR code with $\{n > d + 1, k, d, \alpha, \beta\}$.*

Since they still are MSR codes, they necessarily retain the properties in Section 3.1. Thus, we have

$$\begin{cases} H(S^1(\mathbb{M})) = H(W_1) + (k - 1)H(S_3^1(\mathbb{M})) = \alpha + (k - 1)\beta = d\beta \\ H(S^1(\mathbb{M}')) = H(W_1) + (k - 1)H(S_3^1(\mathbb{M}')) = \alpha + (k - 1)\beta = d\beta, \end{cases} \quad (37)$$

where $H(S^1(\mathbb{M}))$ and $H(S^1(\mathbb{M}'))$ respectively represent the repair data of node 1 under different contexts of truncated MSR codes. Besides, it follows from Lemma 6 that $H(S_{i_1}^1(\mathbb{M})) = H(S_3^1(\mathbb{M}))$ and $H(S_{i_2}^1(\mathbb{M}')) = H(S_3^1(\mathbb{M}'))$ for any $i_1 \in \mathbb{M}$ and $i_2 \in \mathbb{M}'$.

³ It would be reasonable to assume here that the identity of node j can be recognized by eavesdropper, although node j when failed will be replaced by newcomer nodes. In this case, the eavesdropper will gain access to all repair data via sitting on the same node j undergoing different repair epochs.

Furthermore, similar to the deduction of properties given in Section 3.1, we derive

$$\left\{ \begin{aligned} & H(S^1(\mathbb{M}), S^1(\mathbb{M}')) \\ &= H(W_1, S^1(\mathbb{M}), S^1(\mathbb{M}')) \\ &= H(W_1) + H(S^1(\mathbb{M}), S^1(\mathbb{M}')|W_1) \\ &= H(W_1) + H(S^1_{[2,d+1]}(\mathbb{M}), S^1_{[3,d+2]}(\mathbb{M}')|W_1) \\ &= H(W_1) + H(S^1_{[3,k+1]}(\mathbb{M}, \mathbb{M}')|W_1) + H(S^1_{[2,k+2,\dots,d+1]}(\mathbb{M}), S^1_{[k+2,d+2]}(\mathbb{M}')|W_1, S^1_{[3,k+1]}(\mathbb{M}, \mathbb{M}')) \\ &= H(W_1) + H(S^1_{[3,k+1]}(\mathbb{M}, \mathbb{M}')|W_1) \\ &= H(W_1) + H(S^1_{[3,k+1]}(\mathbb{M}, \mathbb{M}')) \\ &= H(W_1) + (k-1)H(S^1_3(\mathbb{M}, \mathbb{M}')), \end{aligned} \right. \quad (38)$$

where $H(S^1_{[2,k+2,\dots,d+1]}(\mathbb{M}), S^1_{[k+2,d+2]}(\mathbb{M}')|W_1, S^1_{[3,k+1]}(\mathbb{M}, \mathbb{M}')) = 0$ results from Lemma 4.

If $S^1_3(\mathbb{M})$ does not share the same information with $S^1_3(\mathbb{M}')$, it has to be that $H(S^1_3(\mathbb{M}, \mathbb{M}')) > \beta$, which leads to that $H(S^1(\mathbb{M}, \mathbb{M}')) > d\beta$. It means that eavesdropper will inevitably obtain different data information after multiple repair epochs with different sets of helper nodes. When traversing all possible truncated MSR codes corresponding to repair epochs with all possible sets of helper nodes, it even may render the storage system unable to maintain any data secrecy.

Based on the above security concern, we define a special MSR code as follows.

Definition 7. (*Stable MSR Code*): A stable MSR code with $\{n > d+1, k, d, \alpha, \beta\}$ is an MSR code with the “stable” repair property, i.e., the data transmitted from any node i to repair node j is independent of the set of helper nodes including the same node i . In other words, content of S^j_i remains invariant under different sets of helper nodes including the same node i .

One can check that the product-matrix-based MSR code [15] is a *stable* MSR code. The following theorem will show that this stable property in fact is the equivalent condition of secrecy capacity between any MSR code with $\{n > d+1, k, d, \alpha, \beta\}$ and its truncated one with $\{n = d+1, k, d, \alpha, \beta\}$.

Lemma 7. Let \mathbb{N} be a stable MSR code with the parameter set $\{n > d+1, k, d, \alpha, \beta\}$ and \mathbb{N}' be the stable MSR code with $\{n = d+1, k, d, \alpha, \beta\}$ truncated from \mathbb{N} , then the secrecy capacity of \mathbb{N} is same as that of \mathbb{N}' .

Proof. Without loss of generality, assume \mathbb{N}' is comprised of the nodes set $[1, d+1]$ truncated from \mathbb{N} . We set the same subsets E, F, G for \mathbb{N} and \mathbb{N}' , where E, F, G are three disjoint subsets of $[1, d+1]$ as defined in section 2.4.

Lemma 3 indicates, for any secure regenerating code with $\{n = d+1, k, d, \alpha, \beta\}$,

$$B^{(s)} \leq H(W_G|W_E, W_F) - H(S^F|W_E, W_F). \quad (39)$$

Although this universal upper bound is established on regenerating codes with length equaling to $\{d+1\}$, it actually is also applicable to those extended regenerating codes with $\{n > d+1\}$ since they still have the reconstruction and regeneration properties. Nevertheless, in order to avoid confusion, we let $S^F(\mathbb{N})$ and $S^F(\mathbb{N}')$ respectively represent the repair data of the nodes set F under the contexts of \mathbb{N} and \mathbb{N}' . Accordingly, we have

$$\left\{ \begin{aligned} & B^{(s)}(\mathbb{N}) \leq H(W_G|W_E, W_F) - H(S^F(\mathbb{N})|W_E, W_F) \\ & B^{(s)}(\mathbb{N}') \leq H(W_G|W_E, W_F) - H(S^F(\mathbb{N}')|W_E, W_F), \end{aligned} \right. \quad (40)$$

with which we only need to prove that $H(S^F(\mathbb{N})|W_E, W_F) = H(S^F(\mathbb{N}')|W_E, W_F)$. Since \mathbb{N} and \mathbb{N}' both are *stable* MSR codes, we can unambiguously substitute $S^F(\mathbb{N}) = S^F_{[1,n]}$ and $S^F(\mathbb{N}') = S^F_{[1,d+1]}$. Thus, it follows by showing that

$$H(S^F_{[1,n]}|W_E, W_F) = H(S^F_{[1,d+1]}|W_E, W_F), \quad (41)$$

with which it is sufficient to prove that $H(S_{[1,n]}^F | S_{[1,d+1]}^F) = 0$. To this end, it is equivalent to prove that $H(S_{[1,n]}^i | S_{[1,d+1]}^i) = H(S_{[d+2,n]}^i | S_{[1,d+1]}^i) = 0$ for any $i \in F$.

Without any loss of generality, we consider the situation when $i = 1$. Thus, we have

$$\begin{cases} H(S_{[d+2,n]}^1 | S_{[1,d+1]}^1) \\ = H(S_{[d+2,n]}^1 | S_{[2,d+1]}^1) \\ = H(S_{[d+2,n]}^1 | W_1, S_{[2,d+1]}^1) \\ \leq H(S_{[d+2,n]}^1 | W_1, S_{[2,k]}^1). \end{cases} \quad (42)$$

1. When $n - d - 1 \geq d - k + 1$, we set Q is any subset of $[d + 2, n]$ of size $d - k + 1$. Because N 's any truncated code with $\{d + 1, k, d, \alpha, \beta\}$ still is an MSR code, the nodes set $\{[2, k] \cup Q\}$ can be viewed as a truncated MSR code. Due to the second term of equation (23) in Lemma 4, we further derive $H(S_Q^1 | W_1, S_{[2,k]}^1) = 0$. Since Q is a random subset of $[d + 2, n]$, it is obvious that $H(S_{[d+2,n]}^1 | W_1, S_{[2,k]}^1) = 0$.

2. When $n - d - 1 < d - k + 1$, we set Q is any $d - k + 1$ -sized set such that $[d + 2, n] \subset Q$ and $[1, k] \cap Q = \emptyset$. Similarly, we have $H(S_Q^1 | W_1, S_{[2,k]}^1) = 0$, from which we can also derive $H(S_{[d+2,n]}^1 | W_1, S_{[2,k]}^1) = 0$.

Combined with formula (42), both cases imply that $H(S_{[d+2,n]}^1 | S_{[1,d+1]}^1) = 0$.

Remark 7 Lemma 7 indicates that secrecy capacity of stable MSR codes does not depend on the parameter n but the remaining parameters $\{k, d, \alpha, \beta, B\}$. One example of stable MSR codes with $\{n > d + 1\}$ is the product-matrix-based MSR code given by Rashmi et al [15]. In another aspect, it is an interesting question to design an MSR code with unstable property. However, for any unstable MSR code with $\{n > d + 1\}$, its secrecy capacity is strictly less than that of the corresponding truncated one with $\{n = d + 1\}$, as shown in Example 1. Thus, this stable property is highly advantageous in constructing secure MSR codes.

Note 2 In subsequent discussion, we focus on the secrecy capacity of linear MSR codes with $\{n = d + 1, k, d, \alpha, \beta\}$.

4 SECRECY CAPACITY OF LINEAR MSR CODES

In this section, we will give a comprehensive and explicit result on secrecy capacity for linear MSR codes with $\{n = d + 1, k, d, \alpha, \beta\}$, which is divided into two categories. In the first category, the secrecy capacity is fully characterized, which applies to all linear scalar MSR codes, i.e., $\beta = 1$. In the second category, upper bounds on secrecy capacity are present, which apply to all known vector codes with $\{\beta = (d - k + 1)^x\}$ where $x \geq 1$ such as Zigzag code [23]. Furthermore, these two categories will be shown to also apply to those unexplored linear vector MSR codes with $\{1 < \beta < d - k + 1\}$. Before these, we first give a lemma that will be used in the subsequent proofs.

Lemma 8. Given any regenerating code with $\{n = d + 1, k, d, \alpha, \beta\}$, for any set $J = (j_1, j_2, \dots, j_m) \subseteq [1, d + 1]$, we have

$$H(S^J) = H(S_{\{[1,d+1] \setminus j_1\}}^{j_1}, S_{\{[1,d+1] \setminus (j_1, j_2)\}}^{j_2}, \dots, S_{\{[1,d+1] \setminus (j_1, j_2, \dots, j_m)\}}^{j_m}). \quad (43)$$

Proof. The proof can be obtained from two directions.

First, it is clear that

$$\begin{cases} H(S_{\{[1,d+1] \setminus j_1\}}^{j_1}, S_{\{[1,d+1] \setminus (j_1, j_2)\}}^{j_2}, \dots, S_{\{[1,d+1] \setminus (j_1, j_2, \dots, j_m)\}}^{j_m}) \\ \leq H(S^J). \end{cases} \quad (44)$$

Second, we can deduce that

$$\begin{cases}
H(S_{\{[1,d+1] \setminus j_1\}}^{j_1}, S_{\{[1,d+1] \setminus (j_1, j_2)\}}^{j_2}, \dots, S_{\{[1,d+1] \setminus (j_1, j_2, \dots, j_m)\}}^{j_m}) \\
= H(\underbrace{S_{\{[1,d+1] \setminus j_1\}}^{j_1}, W_{j_1}}_{}, S_{\{[1,d+1] \setminus (j_1, j_2)\}}^{j_2}, \dots, S_{\{[1,d+1] \setminus (j_1, j_2, \dots, j_m)\}}^{j_m}) \\
= H(\underbrace{S_{\{[1,d+1] \setminus j_1\}}^{j_1}, W_{j_1}, S_{\{j_2, \dots, j_m\}}^{j_2}}_{}, S_{\{[1,d+1] \setminus (j_1, j_2)\}}^{j_2}, \dots, S_{\{[1,d+1] \setminus (j_1, j_2, \dots, j_m)\}}^{j_m}) \\
= H(S_{\{[1,d+1] \setminus j_1\}}^{j_1}, W_{j_1}, S_{\{[1,d+1] \setminus j_2\}}^{j_2}, \dots, S_{\{[1,d+1] \setminus (j_2, \dots, j_m)\}}^{j_m}) \\
\vdots \\
= H(S_{\{[1,d+1] \setminus j_1\}}^{j_1}, W_{j_1}, S_{\{[1,d+1] \setminus j_2\}}^{j_2}, W_{j_2}, \dots, S_{\{[1,d+1] \setminus j_m\}}^{j_m}, W_{j_m}) \\
\geq H(S_{\{[1,d+1] \setminus j_1\}}^{j_1}, S_{\{[1,d+1] \setminus j_2\}}^{j_2}, \dots, S_{\{[1,d+1] \setminus j_m\}}^{j_m}) \\
= H(S^J),
\end{cases} \quad (45)$$

where the formulas in the braces follow from that, for any $l \in [1, m]$,

$$\begin{cases}
H(S_{\{[1,d+1] \setminus j_l\}}^{j_l}) = H(S_{\{[1,d+1] \setminus j_l\}}^{j_l}, W_{j_l}) \\
H(S_{\{j_{l+1}, \dots, j_m\}}^{j_{l+1}}, W_{j_l}) = H(W_{j_l}).
\end{cases} \quad (46)$$

Remark 8 This lemma indicates that there exist much dependence among repair data of multiple sets of nodes. With it, we can reduce the amount of helper nodes for some failed nodes. Thereby, we can derivatively obtain

$$H(S_{\{[1,d+1] \setminus (j_1, j_2, \dots, j_m)\}}^{j_m} | S^{\{j_1, j_2, \dots, j_{m-1}\}}) = H(S^{\{j_1, j_2, \dots, j_m\}}) - H(S^{\{j_1, j_2, \dots, j_{m-1}\}}), \quad (47)$$

which will be used in the proofs later.

4.1 Category 1: Precise Value of Secrecy Capacity

Here, we will give the precise value of secrecy capacity for linear MSR codes with $\{1 \leq \beta < \frac{d-k+1}{l_2-1}\}$ and as a result prove the optimality of the secure product-matrix-based MSR codes given in [31].

4.1.1 The situation when $\beta = 1$.

Theorem 2. In the linear MSR scenario, for any subsets P and T with $\{|P| = k, |T| = d - k + 1, P \cap T = \emptyset\}$, any F such that $F \subseteq T$ and $|F| \leq k - 1$, and arbitrary $i \notin F$, we have $H(S_i^F) = |F|\beta = |F|$ when $\beta = 1$.

Proof. Assume $P = [1, k]$ and $T = [k + 1, d + 1]$. Without loss of generality, also assume $F = [k + 1, k + c]$, where $c \geq 2$ (as it is trivial when $c = 1$). In the linear MSR scenario, Lemma 2 indicates that $H(S_i^A)$ has to be an integer for any subset $A \subseteq F$ and $i \notin A$.

By proof of contradiction, under the condition $\beta = 1$, we assume c is the smallest value satisfying that $H(S_1^{[k+1, k+c-1]}) = H(S_1^{[k+1, k+c]}) = (c - 1)\beta$. Based on Lemma 6, we know that for any $i \notin [k + 1, k + c]$, it must be that $H(S_i^{[k+1, k+c-1]}) = H(S_i^{[k+1, k+c]}) = (c - 1)\beta$, from which we further derive $H(S_i^{k+c} | S_i^{[k+1, k+c-1]}) = 0$. Then, following from Lemma 8, we have that for any $j \in [1, d - k + 1]$,

$$\begin{cases}
H(S^{[k+1, k+j]}) = H(S_{[1, k] \cup [k+2, d+1]}^{k+1}, S_{[1, k] \cup [k+3, d+1]}^{k+2}, \dots, S_{[1, k] \cup [k+j+1, d+1]}^{k+j}) \\
H(S_{[1, k] \cup [k+j+1, d+1]}^{k+j} | S^{[k+1, k+j-1]}) = H(S^{[k+1, k+j]}) - H(S^{[k+1, k+j-1]})
\end{cases} \quad (48)$$

In one way, since we have $H(S_i^{k+c} | S_i^{[k+1, k+c-1]}) = 0$ for any $i \notin [k+1, k+c]$ from the above assumption, we have

$$\begin{cases} H(S_{[1, k] \cup [k+c+1, d+1]}^{k+c} | S^{[k+1, k+c-1]}) \\ = H(S_{[1, k] \cup [k+c+1, d+1]}^{k+c} | S_{[1, k] \cup [k+2, d+1]}^{k+1}, S_{[1, k] \cup [k+3, d+1]}^{k+2}, \dots, S_{[1, k] \cup [k+c, d+1]}^{k+c-1}) \\ \leq H(S_{[1, k] \cup [k+c+1, d+1]}^{k+c} | S_{[1, k] \cup [k+c+1, d+1]}^{[k+1, k+c-1]}) \\ = 0. \end{cases} \quad (49)$$

In another way, we derive

$$\begin{cases} H(S_{[1, k] \cup [k+c+1, d+1]}^{k+c} | S^{[k+1, k+c-1]}) \\ = H(S^{[k+1, k+c]} - H(S^{[k+1, k+c-1]}) \\ = \{H(W_{[k+1, k+c]}) + H(S_{G'}^{[k+1, k+c]})\} - \{H(W_{[k+1, k+c-1]}) + H(S_{G''}^{[k+1, k+c-1]})\} \\ = \{c\alpha + (k-c)(c-1)\beta\} - \{(c-1)\alpha + (k-c+1)(c-1)\beta\} \\ = \alpha - (c-1)\beta \\ = (d-k-c+2)\beta, \end{cases} \quad (50)$$

where

$$\begin{cases} H(S^{[k+1, k+c]}) = H(W_{[k+1, k+c]}) + H(S_{G'}^{[k+1, k+c]}) \\ H(S^{[k+1, k+c-1]}) = H(W_{[k+1, k+c-1]}) + H(S_{G''}^{[k+1, k+c-1]}) \end{cases} \quad (51)$$

result from Lemma 5 and G' , G'' are defined as in Lemma 6 with $|G'| = k-c$ and $|G''| = k-c+1$.

Now, we are to make comparison between equation (49) and (50), when $c \leq \min\{d-k+1, k-1\}$. Equation (50) is a monotone decreasing function in the variable c , thus there are two cases as follows.

1. If $d-k+1 \geq k-1$, when $c = k-1$, equation (50) takes minimum value $\{d-2k+3\}\beta$ that is strictly greater than 0.

2. If $d-k+1 \leq k-1$, when $c = d-k+1$, equation (50) reaches minimum value β that is still positive.

To this end, both cases indicate that equation (50) contradicts formula (49), when $c \leq \min\{d-k+1, k-1\}$, i.e., the assumption that $H(S_i^{[k+1, k+c-1]}) = H(S_i^{[k+1, k+c]})$ cannot hold. In other words, there does not exist such value c that $H(S_i^{[k+1, k+c-1]}) = H(S_i^{[k+1, k+c]})$, when $\beta = 1$ and $c \leq \min\{d-k+1, k-1\}$. Therefore, we can claim that, for any F such that $F \subseteq T$ and $|F| \leq k-1$, $H(S_i^F) = H(S_i^{[k+1, k+c]}) = c\beta$.

Corollary 1. *In the linear MSR scenario, when $\beta = 1$, we have*

$$B^{(s)} = (k-l_1-l_2)(\alpha-l_2\beta), \quad (52)$$

where $l_1 + l_2 \leq k-1$ and $l_2 \leq d-k+1$.

Proof. Remark 6 implies that $B^{(s)} = (k-l_1-l_2)(\alpha-H(S_g^F))$ in the linear MSR scenario, where $g \notin F$. Combining it with Theorem 2, we obtain this corollary.

Corollary 2. *The product-matrix-based secure MSR code given in [31] is optimal for any $l_1 + l_2 \leq k-1$ and $l_2 \leq d-k+1$.*

Proof. First, the product-matrix-based MSR codes constructed in [15] is established on $\beta = 1$ and is a *stable* MSR code as stated in Remark 7. Then, according to the construction of secure MSR codes in [31], the (l_1, l_2) -secure MSR code achieves

$$B^{(s)} = (k-l_1-l_2)(\alpha-l_2\beta). \quad (53)$$

Thus, the secrecy capacity of secure MSR codes in [31] exactly complies with that given in Corollary 1.

Remark 9 Actually, Corollary 1 is applicable to all linear scalar MSR codes, i.e., linear MSR codes with $\beta = 1$. In other words, by MRD code's pre-coding as stated in Remark 2, all linear scalar secure MSR codes can offer this secrecy capacity with precise value given in Corollary 1.

4.1.2 The situations when $1 \leq \beta < \frac{d-k+1}{l_2-1}$.

Theorem 3. In the linear MSR scenario, for any subsets P and T where $\{|P| = k, |T| = d-k+1, P \cap T = \emptyset\}$, any F such that $F \subseteq T$ and $|F| \leq k-1$, and arbitrary $i \notin F$, when $\beta < \frac{d-k+1}{|F|-1}$ or $|F| < 1 + \frac{d-k+1}{\beta}$, we have $H(S_i^F) = |F|\beta$ where $\beta > 1$.

Proof. Similar to Theorem 2, we assume $P = [1, k], T = [k+1, d+1]$ and $F = [k+1, k+c]$ where $c \geq 2$.

By proof of contradiction, we assume c is the smallest value such that $H(S_1^{[k+1, k+c-1]}) = (c-1)\beta$ and $H(S_1^{[k+1, k+c]}) = (c-1)\beta + \theta$, where $\theta \in [0, \beta-1]$ and θ must be an integer following from Lemma 2, when $\beta > 1$. From Lemma 6, we know $H(S_i^{[k+1, k+c-1]}) = (c-1)\beta$ and $H(S_i^{[k+1, k+c]}) = (c-1)\beta + \theta$, where $\theta \in \mathbb{Z} \cap [0, \beta-1]$ for any $i \notin [k+1, k+c]$, from which we further have $H(S_i^{k+c} | S_i^{[k+1, k+c-1]}) = \theta$.

Due to the similar way of Lemma 8 used in the proof of Theorem 2, we first have

$$\begin{cases} H(S_{[1, k] \cup [k+c+1, d+1]}^{k+c} | S^{[k+1, k+c-1]}) \\ = H(S^{[k+1, k+c]}) - H(S^{[k+1, k+c-1]}) \\ = \{H(W_{[k+1, k+c]}) + H(S_{G'}^{[k+1, k+c]})\} - \{H(W_{[k+1, k+c-1]}) + H(S_{G''}^{[k+1, k+c-1]})\} \\ = \{c\alpha + (k-c)[(c-1)\beta + \theta]\} - \{(c-1)\alpha + (k-c+1)(c-1)\beta\} \\ = (d-k-c+2)\beta + (k-c)\theta. \end{cases} \quad (54)$$

In another way, we obtain

$$\begin{cases} H(S_{[1, k] \cup [k+c+1, d+1]}^{k+c} | S^{[k+1, k+c-1]}) \\ = H(S_1^{k+c} | S^{[k+1, k+c-1]}) + H(S_2^{k+c} | S^{[k+1, k+c-1]}, S_1^{k+c}) + \dots + H(S_{d+1}^{k+c} | S^{[k+1, k+c-1]}, S_{[1, k] \cup [k+c+1, d]}^{k+c}) \\ \leq H(S_1^{k+c} | S_1^{[k+1, k+c-1]}) + H(S_2^{k+c} | S_2^{[k+1, k+c-1]}) + \dots + H(S_{d+1}^{k+c} | S_{d+1}^{[k+1, k+c-1]}) \\ = (d-c+1)\theta. \end{cases} \quad (55)$$

Thus, if $(d-c+1)\theta < (d-k-c+2)\beta + (k-c)\theta$, i.e., $(d+1-k)\theta < (d-k-c+2)\beta$, contradiction arises. Particularly, when $\theta = \beta-1$, $(d+1-k)\theta$ reaches maximum $(d+1-k)(\beta-1)$. By simplification, we obtain that, when $\beta < \frac{d-k+1}{c-1}$ or $c < 1 + \frac{d-k+1}{\beta}$, equations (54) contradicts formula (55), which means the assumption that $H(S_1^{[k+1, k+c-1]}) = (c-1)\beta$ and $H(S_1^{[k+1, k+c]}) = (c-1)\beta + \theta$ cannot hold, when $\theta \in \mathbb{Z} \cap [0, \beta-1]$. That is to say, the value of θ here can only be exactly taken by β .

Therefore, for any F such that $F \subseteq T$ and $|F| \leq k-1$, when $\beta < \frac{d-k+1}{c-1}$ or $c < 1 + \frac{d-k+1}{\beta}$, we have $H(S_i^F) = H(S_i^{[k+1, k+c]}) = c\beta$.

Corollary 3. In the linear MSR scenario, when $\beta \geq 1$, we still have

$$B^{(s)} = (k-l_1-l_2)(\alpha-l_2\beta), \quad (56)$$

when $l_1 + l_2 \leq k-1$ and $l_2 < 1 + \frac{d-k+1}{\beta}$ or $\beta < \frac{d-k+1}{l_2-1}$.

Proof. Combining Theorem 2 and Theorem 3, this corollary can be derived as Corollary 1.

Remark 10 In this category, achievability can be attributed to that $H(S_g^F)$ exactly reaches the maximal value $l_2\beta$, when $l_2 < 1 + \frac{d-k+1}{\beta}$. In other words, there does not exist the intersection pattern (dependence) within S_g^F in this category, i.e., all repair data included in S_g^F are mutually independent. However, sometimes $H(S_g^F)$ cannot be exactly calculated and only can be estimated, which will be shown next.

4.2 Category 2: Upper Bounds on Secrecy Capacity

In the other situations when $\beta \geq \frac{d-k+1}{l_2-1}$, we cannot exactly calculate the value of $H(S_g^F)$. Instead, we can only estimate the range of value that $H(S_g^F)$ can be taken from.

4.2.1 The situations when $l_2 = t + 1$, $\frac{d-k+1}{t} \leq \beta < \frac{d-k+1}{t-1}$.

Theorem 4. *Given a linear MSR code, for $l_1 + l_2 \leq k - 1$ and $l_2 = t + 1$, when $\frac{d-k+1}{t} \leq \beta < \frac{d-k+1}{t-1}$, we have*

$$B^{(s)} = (k - l_1 - l_2)(\alpha - \pi(\beta, l_2)), \quad (57)$$

where $\pi(\beta, l_2) = H(S_g^F) \geq t\beta + \frac{d-k-t+1}{d-k+1}\beta$ and $t \leq d - k + 1$.

Proof. It basically follows from formulas (54) and (55) in Theorem 3.

When $\frac{d-k+1}{t} \leq \beta < \frac{d-k+1}{t-1}$, Corollary 3 leads to that $\pi(\beta, t) = t\beta$. According to the proof of Theorem 3, for any $i \notin [k + 1, k + t + 1]$, we have

$$\begin{cases} H(S_i^{[k+1, k+t]}) = t\beta \\ H(S_i^{[k+1, k+t+1]}) = t\beta + \theta, \end{cases} \quad (58)$$

where $\theta \in \mathbb{Z} \cap [0, \beta]$. Because $\beta \geq \frac{d-k+1}{t}$, when setting $\theta = \beta - 1$, we have $(d+1-k)(\beta-1) \geq (d-k-t+1)\beta$, from which we cannot obtain contradiction by formula (54) and (55). With them, we can only derive that $\frac{d-k-t+1}{d-k+1}\beta \leq \theta \leq \beta$. Thus, we obtain

$$\begin{cases} B^{(s)} \\ = (k - l_1 - l_2)(\alpha - \pi(\beta, l_2)) \\ \leq (k - l_1 - l_2)\{\alpha - (t\beta + \frac{d-k-t+1}{d-k+1}\beta)\}, \end{cases} \quad (59)$$

where the equation of the first step follows from Remark 6 and the inequality in the second step results from that $\pi(\beta, l_2) = t\beta + \theta \geq t\beta + \frac{d-k-t+1}{d-k+1}\beta$.

Remark 11 *Our focus in this paper is studying the secrecy capacity of MSR codes that can efficiently repair all nodes under the eavesdropper model with $F \in [1, d+1]$. Unlike Category 1, tightness of the bounds in Theorem 4 stays unclear. In [33], the authors considered using Zigzag code [23] to construct secure MSR code that can attain the upper bound $B^{(s)} \leq (k - l_1 - l_2)(\alpha - (2\beta - \frac{\beta}{n-k}))$, where $\alpha = (n-k)^k$ and $|F| = l_2 = 2$. However, Zigzag code [23] is a systematic MSR code allowing efficient repair of systematic nodes only and the secure Zigzag code designed in [33] is established on the premise that the eavesdropper gains access to the repair data of l_2 systematic nodes, i.e., $F \in [1, k]$.*

Nevertheless, the simple and generally applicable upper bound $B^{(s)} \leq (k - l_1 - l_2)(\alpha - H(S_g^F))$ given in our Theorem 1 in fact also applies to systematic MSR codes, only requiring that $F \in [1, k]$. First, it is clear that the universal upper bound on secrecy capacity for any regenerating code $B^{(s)} \leq H(W_G|W_E, W_F) - H(S^F|W_E, W_F)$ in Lemma 3 is applicable to systematic MSR codes, since they still have the reconstruction property and regeneration property of systematic nodes $[1, k]$. Further due to their minimum storage feature, it can be similarly derived that $H(W_G|W_E, W_F) = H(W_G) = (k - l_1 - l_2)\alpha$. Second, based on the fact that systematic MSR codes have the same parameter setting $\alpha = (d - k + 1)\beta$, one can check that Lemma 4, Lemma 5 and Lemma 6 all apply to systematic MSR codes as well. Thus, Theorem 1 can be applied to systematic MSR codes wherein $F \in [1, k]$.

In the linear MSR scenario, although we assume $F = [k+1, k+l_2]$ in the proof of Theorem 2, Theorem 3 and Theorem 4, they actually all are applicable to linear systematic MSR codes, because there does not restrict F to be necessarily included in $[k + 1, d + 1]$ in their conditions. To this end, systematic MSR codes are supposed to formally share the same secrecy capacity with MSR codes that efficiently repair all

nodes. Consequently, the bound in Theorem 4 also applies to linear systematic MSR codes and actually is consistent with the bound given in [33] for certain situation.

The secure Zigzag code present in [33] is designed by MRD code's pre-coding (Gabidulin code [37]) and is built on $\alpha = (n-k)^k$ and $l_2 = 2$. For Zigzag codes, when a systematic node is failed, the remaining $k-1$ systematic nodes and all the $n-k$ parity nodes are required to participate in repair, which implies that $d = n-1$. Thus, we have that $\alpha = (d-k+1)^k$ and $\beta = (d-k+1)^{k-1} \geq d-k+1$. According to our Theorem 4, we find $t = 1$ satisfies the condition as $\beta = (d-k+1)^{k-1} \geq d-k+1$, which results in that $\pi(\beta, 2) \geq \beta + \frac{d-k}{d-k+1}\beta = 2\beta - \frac{\beta}{d-k+1}$. It exactly equals to $2\beta - \frac{\beta}{n-k}$, the corresponding result of Corollary 16 given in [33]. Furthermore, Corollary 16 in [33] is apparently included in our Theorem 4, since it is not only applicable to the situation $t = 1$.

4.2.2 The situations when $l_2 = t + e, e \geq 1, \frac{d-k+1}{t} \leq \beta < \frac{d-k+1}{t-1}$.

Theorem 5. Given a linear MSR code, for $l_1 + l_2 \leq k-1$ and $l_2 = t + e$, when $\frac{d-k+1}{t} \leq \beta < \frac{d-k+1}{t-1}$, we have

$$B^{(s)} = (k - l_1 - l_2)(\alpha - \pi(\beta, l_2)), \quad (60)$$

where $\pi(\beta, l_2) = H(S_g^F) \geq t\beta + \beta(d-k-t+1)[1 - (\frac{d-k}{d-k+1})^e]$ with $t \leq d-k+1$ and $e \geq 1$.

Proof. Without loss of any generality, we assume the set F is $[1, t+e]$, where $t+e+l_1 \leq k-1$. According to Lemma 6, we know that for any $i \notin [1, t+e]$, $H(S_i^{[1, t+e]})$ is invariant.

Due to $\frac{d-k+1}{t} \leq \beta < \frac{d-k+1}{t-1}$ and Corollary 3, we have

$$\begin{cases} H(S_i^{[1, t+e]}) \\ = t\beta + H(S_i^{t+1}|S_i^{[1, t]}) + \dots + H(S_i^{t+e}|S_i^{[1, t+e-1]}) \\ = t\beta + \theta_1 + \dots + \theta_e, \end{cases} \quad (61)$$

where $H(S_i^{t+j}|S_i^{[1, t+j-1]}) = \theta_j$ and $\theta_j \in \mathbb{Z} \cap [0, \beta]$, for $j \in [1, e]$. Still by Lemma 8, $H(S^{[1, t+e]})$ can be expressed as

$$H(S^{[1, t+e]}) = H(S_{[2, d+1]}^1, S_{[3, d+1]}^2, \dots, S_{[t+e+1, d+1]}^{t+e}). \quad (62)$$

First, with the method similar to the proof of Theorem 2 and 3, we have

$$\begin{cases} H(S_{[t+e+1, d+1]}^{t+e}|S^{[1, t+e-1]}) \\ = H(S^{[1, t+e]}) - H(S^{[1, t+e-1]}) \\ = \{(t+e)\alpha + (k-t-e)[t\beta + \theta_1 + \dots + \theta_e]\} - \{(t+e-1)\alpha + (k-t-e+1)[t\beta + \theta_1 + \dots + \theta_{e-1}]\} \\ = \alpha - [t\beta + \theta_1 + \dots + \theta_{e-1}] + (k-t-e)\theta_e. \end{cases} \quad (63)$$

Second, we obtain

$$\begin{cases} H(S_{[t+e+1, d+1]}^{t+e}|S^{[1, t+e-1]}) \\ \leq (d-t-e+1)\theta_e, \end{cases} \quad (64)$$

which can be derived as inequality (55).

Then, combining equation (63) with (64), we derive $(d-k+1)\theta_e \geq \alpha - [t\beta + \theta_1 + \dots + \theta_{e-1}]$, from which we further have

$$\theta_e \geq \frac{(d-k-t+1)\beta}{d-k+1} - \frac{\theta_1 + \dots + \theta_{e-1}}{d-k+1}. \quad (65)$$

Through rearrangement, it can be changed to

$$\theta_1 + \dots + \theta_e \geq \frac{(d-k-t+1)\beta}{d-k+1} + \frac{(d-k)(\theta_1 + \dots + \theta_{e-1})}{d-k+1}. \quad (66)$$

By setting $\omega(e) = \theta_1 + \dots + \theta_e$, we obtain

$$\omega(e) \geq \frac{d-k}{d-k+1}\omega(e-1) + \frac{(d-k-t+1)\beta}{d-k+1}. \quad (67)$$

From Theorem 4, we know $\theta_1 \geq \frac{d-k-t+1}{d-k+1}\beta$. Hence, by the method of recursion and induction, we have

$$\omega(e) \geq \beta(d-k-t+1)\left[1 - \left(\frac{d-k}{d-k+1}\right)^e\right]. \quad (68)$$

To this end, we have $\pi(\beta, l_2) = H(S_i^{[1, t+e]}) = t\beta + \omega(e) \geq t\beta + \beta(d-k-t+1)\left[1 - \left(\frac{d-k}{d-k+1}\right)^e\right]$.

Remark 12 Theorem 5 is the supplementary of Theorem 4, which expands the range of values that l_2 can be taken from. In Theorem 4, e only can be taken by 1, while Theorem 5 takes e by any value only needing to satisfy $l_1 + t + e \leq k-1$ and $t \leq d-k+1$. As stated in Remark 11, Theorem 5 basically also applies to systematic MSR codes for $F \in [1, k]$.

In fact, the upper bound given in [32] is also a special case of our Theorem 5. Zigzag codes [23] by pre-coding of MRD codes are shown to be able to achieve this bound on secrecy capacity in [32]. Since $\alpha = (d+1-k)^k$, we know $\beta = (d+1-k)^{k-1} > d+1-k$, which similarly indicates that only $t = 1$ conforms the constraints on β required by our Theorem 5. Thereby, we have $l_2 = e + 1$, from which we derive $\pi(\beta, l_2) \geq \beta + \beta(d-k)\left[1 - \left(\frac{d-k}{d-k+1}\right)^{l_2-1}\right]$. By simplification, we further have

$$\begin{cases} \pi(\beta, l_2) \\ \geq \beta + \beta(d-k)\left[1 - \left(\frac{d-k}{d-k+1}\right)^{l_2-1}\right] \\ = \alpha - \beta(d-k)\left(\frac{d-k}{d-k+1}\right)^{l_2-1} \\ = \alpha - \alpha\left(1 - \frac{1}{d-k+1}\right)^{l_2}, \end{cases} \quad (69)$$

which leads to that $B^{(s)} \leq (k-l_1-l_2)\left(1 - \frac{1}{d-k+1}\right)^{l_2}\alpha$. It is exactly consistent with the bound given in [32], i.e., $B^{(s)} \leq (k-l_1-l_2)\left(1 - \frac{1}{n-k}\right)^{l_2}\alpha$.

Although some upper bounds (limited to $t = 1$ or $\beta > d-k+1$) in this category are achievable for the Zigzag codes considered in [32,33], they are not generally achievable for all other MSR codes with $\{\beta > d-k+1\}$. For example, for those vector MSR codes with $\{\beta > d-k+1\}$ designed by concatenating m same scalar MSR codes where $m > d-k+1$, their secrecy capacity is exactly equal to $(k-l_1-l_2)(\alpha - l_2\beta)$ following from Corollary 1 where $\beta = m$, since each scalar MSR code within a vector MSR code shares the same code construction and can be designed to be mutually independent. It is obvious⁴ that $(k-l_1-l_2)(\alpha - l_2\beta) < (k-l_1-l_2)\left(1 - \frac{1}{d-k+1}\right)^{l_2}\alpha$, which means that those vector MSR codes cannot reach the bounds in Theorem 5. Therefore, unlike Category 1, the value of $H(S_g^F)$ or $\pi(\beta, l_2)$ in Category 2 cannot be determined, i.e. its precise value may vary with different MSR codes' constructions.

4.3 Putting Together

Now combining the two categorizes on secrecy capacity of linear MSR codes, we give the following comprehensive and explicit result on secrecy capacity for any linear MSR codes with $\{n = d+1\}$.

⁴ In the category where $l_2 = t+e$ and $\frac{d-k+1}{t} \leq \beta < \frac{d-k+1}{t-1}$, we have $B^{(s)} \leq (k-l_1-l_2)\left\{\alpha - t\beta - \beta(d-k-t+1)\left[1 - \left(\frac{d-k}{d-k+1}\right)^e\right]\right\}$. Through analysis, one can check the term $\beta(d-k-t+1)\left[1 - \left(\frac{d-k}{d-k+1}\right)^e\right] < \beta(d-k-t+1)\left[e\left(1 - \frac{d-k}{d-k+1}\right)\right] = e\beta\left(1 - \frac{t}{d-k+1}\right) = e\beta - \frac{et\beta}{d-k+1} \leq e\beta - e < e\beta$. So, we derive $t\beta + \beta(d-k-t+1)\left[1 - \left(\frac{d-k}{d-k+1}\right)^e\right] < (t+e)\beta = l_2\beta$, which leads to that $(k-l_1-l_2)(\alpha - l_2\beta) < (k-l_1-l_2)\left(1 - \frac{1}{d-k+1}\right)^{l_2}\alpha$ when $t = 1$.

Theorem 6. Given a linear MSR code with $\{n = d + 1, k, d, \alpha, \beta\}$, for $l_1 + l_2 \leq k - 1$, we have

$$B^{(s)} = (k - l_1 - l_2)(\alpha - \pi(\beta, l_2)), \quad (70)$$

where $\pi(\beta, l_2)$

$$\begin{cases} = l_2\beta, & \text{if } l_2 \leq t, \beta < \frac{d-k+1}{t-1}; \\ \geq t\beta + \beta(d - k - t + 1)[1 - (\frac{d-k}{d-k+1})^e], & \text{if } l_2 = t + e, \frac{d-k+1}{t} \leq \beta < \frac{d-k+1}{t-1}, \end{cases} \quad (71)$$

where $1 \leq t \leq d - k + 1$ and $e \geq 1$.

Remark 13 In the literature, the known linear MSR codes are comprised of the scalar MSR codes with $\{\beta = 1\}$ [15,16,17,18,19,20] and the vector MSR codes with $\{\beta = (d - k + 1)^x\}$ where $x \geq 1$ [21,22,23,24,25,26,27,28,29]. It should be noted that these vector MSR codes are not designed from concatenation of scalar MSR codes. Similar to Zigzag codes [23], they share the same intersection pattern, i.e. there exist the same dependence within disjoint sets of repair data transmitted from any one node (e.g. S_g^F where $g \notin F$). Thus, the second item in formula (71) also applies to them.

4.4 Further Discussions

Theorem 6 exhibits a comprehensive and explicit result on secrecy capacity for any linear MSR code given the parameter set $\{n = d + 1, k, d, \alpha, \beta\}$ and the (l_1, l_2) -eavesdropper model. In retrospect, all constructions of linear MSR codes are based on the scalar case $\beta = 1$ or partial vector cases where β is required to be exponential in $d - k + 1$. Thus, designing linear vector MSR codes with $\{1 < \beta < d - k + 1\}$ by no concatenation remains open. Nevertheless, our Theorem 6 still presents certain results on secrecy capacity for these unexplored MSR codes. Thereby, we put forward two questions as follows.

Question 1. Do there exist the linear MSR codes with $\{1 < \beta < d - k + 1\}$ by no concatenation? If so, how can we construct them?

Question 2. Given such a construction with $\{1 < \beta < d - k + 1\}$, is it achievable for the bounds given in formula (71) when $l_2 \geq 1 + \frac{d-k+1}{\beta}$?

Remark 14 According to formula (71), for the linear MSR codes with $\{1 < \beta < d - k + 1\}$ by no concatenation, when $l_2 < 1 + \frac{d-k+1}{\beta}$, it must be that

$$B^{(s)} = (k - l_1 - l_2)(\alpha - l_2\beta). \quad (72)$$

However, when $l_2 \geq 1 + \frac{d-k+1}{\beta}$, formula (71) leads to that

$$B^{(s)} \leq (k - l_1 - l_2)(\alpha - t\beta - \beta(d - k - t + 1)[1 - (\frac{d-k}{d-k+1})^{l_2-t}]), \quad (73)$$

where $t = \frac{d-k+1}{\beta}$ if β divides $d - k + 1$, and $t = \lfloor 1 + \frac{d-k+1}{\beta} \rfloor$ if β does not divide $d - k + 1$. Hence, as in Question 2, we ask whether it is achievable for the upper bound (73) given such a code.

Overall, Theorem 6 predicts certain results on secrecy capacity for these unexplored MSR codes, which consist of the precise value of secrecy capacity (72) and the upper bound on secrecy capacity (73).

5 CONCLUSION

In this paper, we carry out research on the secrecy capacity of MSR codes. We assume the passive adversarial model where the eavesdropper can observe the contents of certain nodes and the repair data of some other nodes. Although the secrecy capacity of MBR codes has been characterized completely [30], it is a challenging task to analyze the secrecy capacity of MSR codes [31,32,33]. The additional difficulty

comes from the fact that the amount of data transmitted for a failed node in MSR codes, is not entirely stored on the node undergoing repair, making it challenging to compute the joint entropy of the repair data. With such a system model, we focus on investigating the repair data in the MSR scenario from the information-theoretic perspective.

We first obtain some information-theoretic properties and some upper bounds on secrecy capacity for general MSR codes, in addition to which we introduce a new concept named by *stable* MSR codes. For the unstable MSR codes, we assume the eavesdropper could identify the nodes with repair data captured and demonstrate that its secrecy capacity is strictly less than that of *stable* MSR code. In the linear MSR scenario, we utilize permutation polynomial and orthogonal system in finite fields to explain the fact that entropy of multiple sets of repair data is an integer and ultimately derive a comprehensive and explicit result on secrecy capacity which closely depends on the value of β . This outcome not only explains and generalizes the previous results in [31,32,33], but also predicts certain results for some unexplored linear MSR codes. After that, we put forward two related questions. On the other hand, we find that all of these results also apply to systematic MSR codes with repair data of systematic nodes captured.

References

1. T. M. Cover and J. A. Thomas, "Entropy, Relative Entropy and Mutual Information," Elements of Information Theory, pp. 12–49, 1991.
2. R. Lidl, H. Niederreiter and P. M. Cohn, "Encyclopedia of Mathematics and Its Applications, Finite Fields," Cambridge University Press.
3. H. Delfs and H. Knebl, "Introduction to Cryptography: Principles and Applications," 2nd ed., Springer, 2007.
4. R. Ahlswede, N. Cai, S. Y. R. Li, and R.W. Yeung, "Network Information Flow," IEEE Trans. Inf. Theory, 46(4), pp. 1204–1216, Jul. 2000.
5. C. Huang, H. Simitci, Y. Xu, A. Ogus, B. Calder, P. Gopalan, J. Li, and S. Yekhanin, "Erasure Coding in Windows Azure Storage," Proc. USENIX Annual Technical Conference (ATC), Boston, MA, 2012.
6. H. Weatherspoon and J. D. Kubiatowicz, "Erasure Coding vs. Replication: A Quantitative Comparison," Proc. Int. Workshop. Peer-to-Peer Syst, 2002.
7. A. G. Dimakis, P. B. Godfrey, Y. Wu, M. J. Wainwright, and K. Ramchandran, "Network Coding for Distributed Storage Systems," IEEE Trans. Inf. Theory, 56(9), pp. 4539–4551, 2010.
8. A. G. Dimakis, K. Ramchandran, Y. Wu and C. Suh, "A Survey on Network Codes for Distributed Storage," Proc. IEEE, 99(3), pp. 476–489, 2011.
9. N. B. Shah, K. V. Rashmi, and P. V. Kumar, and K. Ramchandran, "Distributed Storage Codes with Repair-by-Transfer and Nonachievability of Interior Points on The Storage-Bandwidth Tradeoff," IEEE Trans. Inf. Theory, 58(3), pp. 1837–1852, 2012.
10. Iwan M. Duursma, "Outer Bounds for Exact Repair Codes," arXiv preprint arXiv:1406.4852, 2014.
11. Iwan M. Duursma, "Shortened Regenerating Codes," arXiv preprint arXiv:1505.00178, 2015.
12. C. Tian, V. Aggarwal, and V. A. Vaishampayan, "Exact-Repair Regenerating Codes via Layered Erasure Correction and Block Designs," Proc. IEEE Int. Symp. Inf. Theory (ISIT), pp. 1431–1435, Jul. 2013.
13. T. Ernvall, "Exact-Regenerating Codes Between MBR and MSR Points [Online]," Apr. 2013, Available: <http://arxiv.org/abs/1304.5357>.
14. V. R. Cadambe and S. A. Jafar, "Interference Alignment and The Degree of Freedom for The K User Interference Channel," IEEE Trans. Inf. Theory, 54(8), pp. 3425–3441, Aug. 2008.
15. K. V. Rashmi, N. B. Shah, and P. V. Kumar, "Optimal Exact-Regenerating Codes for Distributed Storage at The MSR and MBR Points via A Product-Matrix Construction," IEEE Trans. Inf. Theory, 57(8), pp. 5227–5239, Aug. 2011.
16. C. Suh and K. Ramchandran, "Exact-Repair MDS Codes for Distributed Storage Using Interference Alignment," Proc. IEEE International Symposium on Information Theory (ISIT), Austin, pp. 161–165, Jun. 2010.
17. Y. Wu and A. G. Dimakis, "Reducing Repair Traffic for Erasure Coding-Based Storage via Interference Alignment," Proc. IEEE Int. Symp. Inf. Theory, Seoul, Korea, pp. 2276–2280, Jul. 2009.
18. K. V. Rashmi, N. B. Shah, P. V. Kumar, and K. Ramchandran, "Explicit Construction of Optimal Exact Regenerating Codes for Distributed Storage," Proc. 47th Annual Allerton Conference on Communication, Control, and Computing, Urbana-Champaign, pp. 1243–1249, Sep. 2009.
19. K. V. Rashmi, N. B. Shah, and P. V. Kumar, "Regenerating Codes for Errors and Erasures in Distributed Storage," Proc. IEEE International Symposium on Information Theory (ISIT), Cambridge, MA, 2012.

20. N. B. Shah, K. V. Rashmi, P. V. Kumar, and K. Ramchandran, "Interference Alignment in Regenerating Codes for Distributed Storage: Necessity and Code Constructions," *IEEE Trans. Inf. Theory*, 56(4), pp. 2134–2158, 2012.
21. Z. Wang, I. Tamo, and J. Bruck, "On Codes for Optimal Rebuilding Access," In *Communication, Control, and Computing (Allerton)*, 2011 49th Annual Allerton Conference on. IEEE, pp. 1374–1381, 2011.
22. D. S. Papailiopoulos, A.G. Dimakis, and V. R. Cadambe, "Repair Optimal Erasure Codes through Hadamard Designs," In *Allerton Conference on Control, Computing, and Communication*, Urbana-Champaign, IL, pp. 1382–1389, 2011.
23. I. Tamo, Z. Wang, and J. Bruck, "Zigzag Codes: MDS Array Codes with Optimal Rebuilding," *IEEE Trans. Inf. Theory*, 59, pp. 1597–1616, march. 2013.
24. Z. Wang, I. Tamo, and J. Bruck, "Long MDS Codes for Optimal Repair Bandwidth," *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 1182–1186, Jul. 2012.
25. V. R. Cadambe, C. Huang, S. A. Jafar, and J. Li, "Optimal Repair of MDS Codes in Distributed Storage via Subspace Interference Alignment," *Tech. Rep. arXiv:1106.1250*, 2011.
26. V. R. Cadambe, C. Huang, J. Li, and S. Mehrotra, "Polynomial Length MDS Codes with Optimal Repair in Distributed Storage," In *Signals, Systems and Computers (ASIOMAR)*, 2011 Conference Record of the Forty Fifth Asilomar Conference on. IEEE, pp. 1850–1854, 2011.
27. G. K. Agarwal, B. Sasidharan, and P. V. Kumar, "An Alternate Construction of An Access-Optimal Regenerating Code with Optimal Subpacketization Level," In *National Conference on Communication (NCC)*, 2015.
28. Y. S. Han, H. T. Pai, R. Zheng, and P. K. Varshney, "Update-Efficient Regenerating Codes with Minimum Per-Node Storage," *Proc. IEEE International Symposium on Information Theory (ISIT)*, pp. 1436–1440, 2013.
29. J. Li, X. Tang, and U. Parampalli, "A Framework of Constructions of Minimal Storage Regenerating Codes with the Optimal Access/Update Property," *arXiv preprint arXiv:1311.4947*, 2013.
30. S. Pawar, S. El. Rouayheb, and K. Ramchandran, "Securing Dynamic Distributed Storage Systems Against Eavesdropping and Adversarial Attacks," *IEEE Trans. Inf. Theory*, 57(10), pp. 6734–6753, Oct. 2011.
31. N. B. Shah, K. V. Rashmi, and P. V. Kumar, "Information-Theoretically Secure Regenerating Codes for Distributed Storage," *Proc. IEEE Globecom*, Houston, USA, pp. 1–5, Dec. 2011.
32. S. Goparaju, S. El. Rouayheb, R. Calderbank, and H. V. Poor, "Data Secrecy in Distributed Storage Systems under Exact Repair," *Proc. Symp. Netw. Coding*, pp. 1–6, 2013.
33. A. S. Rawat, O. O. Koyluoglu, N. Silberstein, and S. Vishwanath, "Optimal Locally Repairable and Secure Codes for Distributed Storage Systems," *IEEE Trans. Inf. Theory*, 60(1), pp. 212–236, 2014.
34. R. Tandon, S. Amuru, T. C. Clancy, and R. M. Buehrer, "Towards Optimal Secure Distributed Storage Systems with Exact Repair," *IEEE Trans. Inf. Theory*, 60(6), pp. 3477–3492, 2016.
35. O. O. Koyluoglu, A. S. Rawat, and S. Vishwanath, "Secure Cooperative Regenerating Codes for Distributed Storage Systems," *IEEE Trans. Inf. Theory*, 60(9), pp. 5228–5244, 2014.
36. R. M. Roth, "Maximum-Rank Array Codes and Their Application to Crisscross Error Correction," *IEEE Trans. Inf. Theory*, 37(2), pp. 328–336, 1991.
37. E. M. Gabidulin, "Theory of Codes with Maximum Rank Distance," *Problems of Information Transmission*, vol. 21, pp. 1–12, July. 1985.
38. N. Prakash, G. M. Kamath, V. Lalitha, and P. V. Kumar, "Optimal Linear Codes with A Local-Error-Correction Property," *Proc. IEEE International Symposium on Information Theory (ISIT)*, pp. 2776–2780, 2012.
39. P. Gopalan, C. Huang, H. Simitci, and S. Yekhanin, "On The Locality of Codeword Symbols," *IEEE Trans. Inf. Theory*, 58(11), pp. 6925–6934, 2012.
40. D. S. Papailiopoulos and A. G. Dimakis, "Locally Repairable Codes," *Proc. IEEE International Symposium on Information Theory (ISIT)*, pp. 2771–2775.